BIOMETRICAL SPEAKER DESCRIPTION FROM VOCAL CORD PARAMETERIZATION

P. Gómez, V. Rodellar, A. Álvarez, J. C. Lázaro, K. Murphy, F. Díaz, R. Fernández

Facultad de Informática, Universidad Politécnica de Madrid, Campus de Montegancedo s/n, 28660
Boadilla del Monte, Madrid Spain, phone: +34.91.336.7384, e-mail: pedro@pino.datsi.fi.upm.es

ABSTRACT

The parameterization of vocal cord dynamics may be useful in different fields of voice processing, as in speech synthesis, voice pathology detection, or speaker identification, among others. A technique to determine biomechanical cord parameters is presented to produce estimates of mass, stiffness and losses from the body and cover structures of the vocal cords. Results from a database of normophonic subjects are in good agreement with values used in cord modelling. This technique provides biometrical descriptors of speakers with use in different fields of application.

1. INTRODUCTION

The parameterization of voice is useful in many different fields, as voice pathology research [8][2], vocal tract and cord modeling [1], speech synthesis [7], speaker biometrical description and identification, etc. Parameters obtained directly from voice are subject to contamination from the influence of the vocal tract, and as such, they may not be the best choice in voice pathology detection or in the biometrical characterization of the glottal features of the speaker. For this reason other estimation methods could be of interest, as those based in the biomechanics of the vocal cord. In preliminary work [6] it has been shown that the biomechanics of the vocal cords (masses, springs and losses) is directly related with the spectral distribution of glottal dynamics correlates, as the *glottal source*. Under certain conditions it would be possible to obtain direct estimates of these parameters assuming that the spectral distributions of glottal correlates are directly related to certain functions of cord dynamics [4]. Parameter unbalance between cords, and their deviations with respect to a general population may be related with voice pathologies [5]. Through the present paper a description of the vocal cord into its body and cover structures, and estimates of their respective biomechanical parameters are provided. Results from a database including adult and young individuals from both sexes are given. The estimates are compared with values used in vocal cord modeling.

2. VOCAL CORD MODELING

The present study is based in the well-known *3-mass* model of Story and Titze [10] shown in Figure 1.a, where the vocal cord body (integrated by muscles responsible of cord tension) is represented by masses M_{bl} and M_{br}, for the left and right vocal cords, and by springs K_{bl} and K_{br} to account for the elastic behavior of the bulk tissues, linking body masses to the walls of the thyro-arytenoid cartilages.

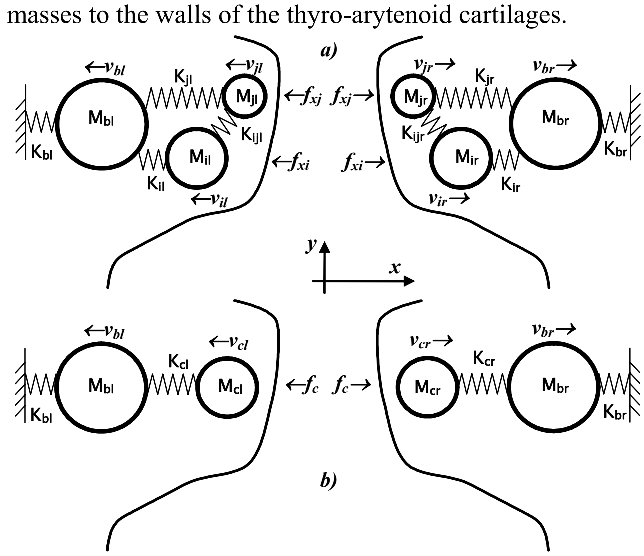

Figure 1. a) *3-mass* model of the vocal cords showing the parameters and dynamic variables involved. b) *2-mass* model when the cover masses are rigidly linked.

The cord cover (integrated by a three-layered epithelial structure) is represented by a pair of masses per cord $M_{il,r}$ and $M_{jl,r}$ linked to the body mass by springs $K_{il,r}$ and $K_{jl,r}$, and intercoupled by springs $K_{ijl,r}$. Masses and springs are lumped parameters, associated to point-like structures. The system is only allowed to move along the *x* axis, $f_{xi,j}$ being forces and $v_{il,r}$, $v_{jl,r}$ and $v_{bl,r}$ being the respective mass velocities. If the study is to stress the role of the cover structure, highlighting the phenomenon known as the *mucosal wave*, the coupling springs between cover masses are assigned small values, to allow both cover masses to vibrate somewhat loosen. As the present study wants to stress the relative role between body and cover it will be assumed that coupling springs are rigid bars such that the whole structure could be simplified to a simpler *2-mass* model as the one shown in Figure 1.b, where

$$M_{cl,r} = M_{il,r} + M_{jl,r}$$
$$K_{cl,r} = K_{il,r} + K_{jl,r} \tag{1}$$

In this way the study of vocal cord dynamics can be divided into that of the cord body and the cover structures. The processing methodology is based in the separation of vocal cord vibration correlates (*glottal source*, obtained from the inversion of the vocal tract [1]) into two components: a phonation-cycle long average signal known as the *average acoustic wave*, and a component keeping higher frequency contents of vibration, which could be referred as the *cover dynamics component* or the *mucosal wave correlate*.

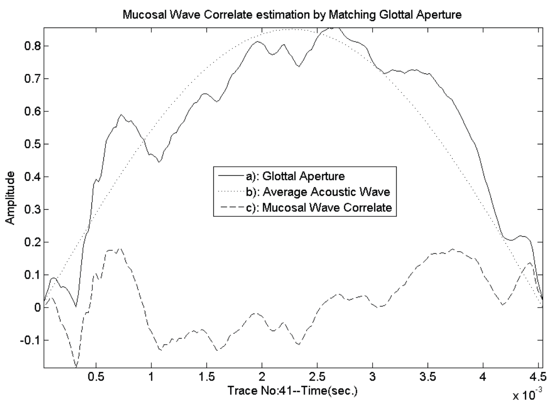

Figure 2. Splitting the *glottal source* (in full line) into the *average acoustic wave* (dot) and the *cover dynamics component* (dash).

The *glottal source* can be related with the *glottal aperture*, considered as the average distance between vocal cord average cover masses during phonation, and the *cover dynamics component* associated with the relative movement of cover masses $M_{cl,r}$ respect to body masses $M_{bl,r}$ (Figure 1.b). Different separation algorithms can be used [3], low-pass filtering being a reasonable choice. Both components of the *glottal source* can be seen in Figure 2 after separation.

3. VOCAL CORD PARAMETER ESTIMATION

To determine the parameters of the the body and cover dynamics the electromechanical equivalent represented by the system given in Figure 3 is used.

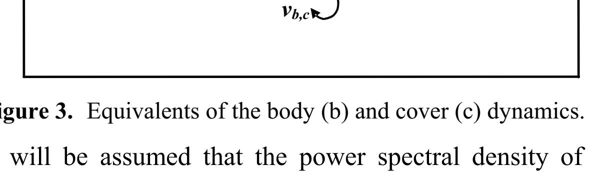

Figure 3. Equivalents of the body (b) and cover (c) dynamics.

It will be assumed that the power spectral density of the *glottal source components* will be related (within a scale factor) with the square modulus of the input admittance of each electromechanical equivalent. The estimation of the body parameters from the *average acoustic wave* is not complicate, because its *power spectral density* is smooth and predictable, as most of the variability of the *glottal source* is present in the *cover dynamics component*. The process of estimation is as follows

$$M_{bl,r} = \frac{\omega_2}{\omega_2^2 - \omega_r^2}\left[\frac{T_r - T_2}{T_r T_2}\right]^{1/2} \tag{2}$$

ω_r being the resonance frequency given by

$$\omega_r^2 = \frac{K_{bl,r}}{M_{bl,r}} \tag{3}$$

where the square modulus of the admittance is given by

$$T(\omega) = \frac{1}{\left[R_{bl,r}^2 + \varpi^2 M_{bl,r}^2\right]^2} \tag{4}$$

with the frequency relative to the resonance point

$$\varpi = \frac{\omega^2 - \omega_r^2}{\omega} \tag{5}$$

and

$$T_r = T(\omega = \omega_r) = \frac{1}{R_{bl,r}^2}$$
$$T_2 = T(\omega = 2\omega_r) \tag{6}$$

The estimation procedure must detect the value of pitch, which is used to evaluate ω_r. The determination of T_r and T_2 is carried out on the power spectral density of the *average acoustic wave*. This leads to the determination of the losses from (6) and to the mass (2) and stiffness (3). The estimation results cycle by cycle from a voice trace *0.2 sec* long *{198*, see Table 1*}* are plotted in Figure 5 (templates 35-37), together with their statistical spread to evaluate the reliability of the estimates.

3.1. Adaptive estimation of cover parameters

The estimation of the parameters from the *cover dynamics component* is carried out in two steps. First a rough estimation based in (2)-(6) is carried out. As the power spectral density of the *cover dynamics component* is more irregular, this leading to less robust estimations of the pairs $[\omega_r, T_r]$ and $[\omega_n, T_n]$, an adaptive refinement is implemented. For such the following cost function is defined

$$L = \int_0^{\omega_s} \left[T_u(\omega) - T_a(\omega)\right]^2 d\omega \tag{7}$$

where ω_s is the *Nyquist frequency*, T_u is the power spectral density of the *cover dynamics component*, and T_a is the adaptive evaluation of (4). This function is known to have a single minimum, which may be obtained either by a block expression (which is not straight forward) or through step-wise estimations using the gradient descent as

$$M_{bl,r}^i = M_{bl,r}^{i-1}\left[1 - 4\int_0^{\omega_s}(T_u - T_a)T_a^2\varpi^2 d\omega\right] \tag{8}$$

i being the step index, which is repeated to reduce (7) to a value as small as desired. The results of adaptation for the same voice trace may be seen in Figure 4.

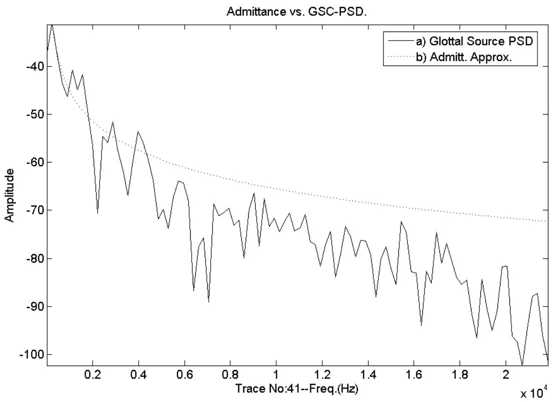

Figure 4. Adjusting the admittance of the *1-mass* equivalent model (dot) against the power spectral density of the *cover dynamics component* (full) by the adaptive evaluation of mass.

The results of estimating the cover parameters on a phonation cycle basis may be seen in Figure 5 (templates 41-43). The refinement in the estimation obtained by adaptation may be of about 20-50% that of total estimation.

4. RESULTS AND DISCUSSION

A balanced database of *100 normophonic subjects* was recorded, half of the subjects were adults (age 18-50), half were young (age 12-18), equally distributed by sexes. Normophony was determined from electroglottographic and endoscopic examination and by subjective evaluation. The results presented here correspond to a segment of *24 subjects* randomly selected, including *8 male and 8 female adults*, and *8 non-adults*. The characteristics of each *subject* are given in Table 1. Recordings included three realizations of the vowel */a/* about *3 sec.* long, of which *0.2 sec.* frames were used for the analysis. The frames were processed to extract the *average acoustic wave* and the *cover dynamic component*, from which the *body* and *cover biomechanics* were estimated as described for each phonation cycle. Average values were taken for comparison if the statistical spread was reasonably small, as given in Figure 5 (right). The results for the *24* speakers selected from the database are shown in Table 1. It may be seen from the biomechanical estimates that the physiological differences in the dynamic mass due to sex or age are not relevant, the stiffness appearing as the responsible factor for different resonant frequencies and therefore, for pitch. In some cases the dynamic mass estimates for the cover are larger than those for the body, as in samples *{191, 197, 198, 1a5, 1ae, 1b1}*. In all the cases studied the losses associated to the cover are larger than those of the body. This observation could be justified by the fact that most of turbulence and heating takes place at and around the cover. An underlying problem is that of the uniqueness of mass and stiffness estimates, as the pair $K/M = \omega_r^2$ may accept infinite pairs of values as far as their quotient remains unaltered.

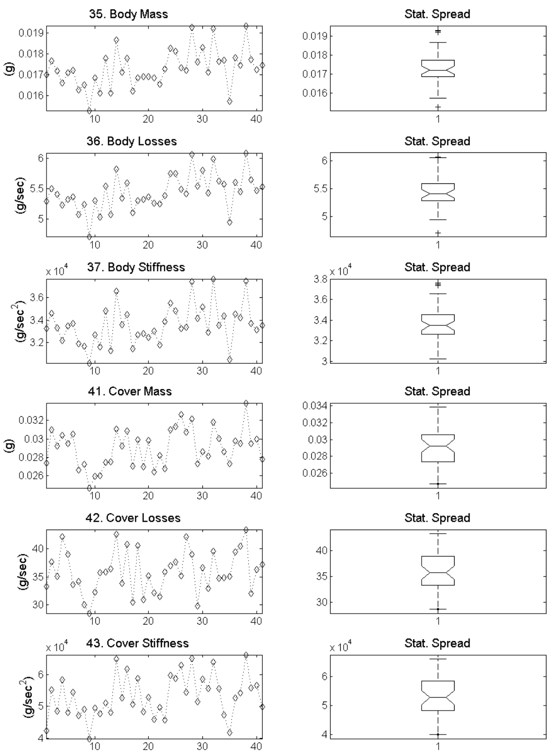

Figure 5. Left: Body (35-37) and cover (41-43) estimates for each phonation cycle (x axis). Right: respective statistical distributions.

T\P	Char	M_b	R_b	K_b	M_c	R_c	K_c
191	YF	0.00209	0.626	3.78×10^3	0.00220	2.24	3.24×10^3
192	AM	0.00803	1.160	3.33×10^3	0.00610	4.26	2.27×10^3
193	YF	0.00521	1.370	7.57×10^3	0.00477	6.50	7.80×10^3
194	YM	0.01040	2.230	9.53×10^3	0.00797	8.36	8.54×10^3
195	YM	0.00563	1.380	6.73×10^3	0.00367	4.42	4.69×10^3
196	AM	0.00463	0.807	2.80×10^3	0.00290	2.49	2.10×10^3
197	YF	0.00396	1.270	8.01×10^3	0.00560	7.00	1.50×10^4
198	YM	0.01730	5.440	3.36×10^4	0.02900	35.9	5.33×10^4
199	AM	0.00584	0.855	2.52×10^3	0.00445	3.38	3.92×10^3
19A	YF	0.00451	1.180	6.74×10^3	0.00406	6.28	1.16×10^4
19B	AM	0.01180	2.100	7.40×10^3	0.01060	8.61	9.59×10^3
19C	AM	0.00950	1.410	4.10×10^3	0.00695	4.65	2.52×10^3
19D	YM	0.00910	2.990	1.91×10^4	0.00811	10.6	1.49×10^4
19E	AM	0.00542	0.852	2.80×10^3	0.00381	2.77	1.70×10^3
19F	AM	0.00674	1.180	4.11×10^3	0.00526	4.81	4.52×10^3
1A0	AM	0.00543	0.839	2.49×10^3	0.00362	2.39	1.25×10^3
1A5	AF	0.00729	1.950	1.06×10^4	0.00793	7.68	8.73×10^3
1AA	AF	0.00499	1.380	8.01×10^3	0.00400	6.11	1.11×10^4
1AB	AF	0.00523	1.440	7.88×10^3	0.00465	5.61	7.29×10^3
1AE	AF	0.00513	1.580	9.82×10^3	0.00766	9.84	1.46×10^4
1B1	AF	0.00406	1.180	6.73×10^3	0.00591	7.19	8.05×10^3
1B9	AF	0.00420	1.230	7.18×10^3	0.00396	5.30	7.21×10^3
1BF	AF	0.00436	1.190	6.40×10^3	0.00390	3.96	4.58×10^3
1C6	AF	0.00419	1.140	6.19×10^3	0.00227	2.94	2.73×10^3

Table 1. Body and cover biomechanical parameters for a database of 24 normophonic samples. Subject characteristics: AM – adult male, AF – Adult female, YM – young male, YF – young female

The factor of losses is nevertheless a strict reference, as it affects the bandwidth of the systems, measured by the height of the resonant peak relative to its width. Having this into account the normalization of power spectral densities on each phonation cycle in reference to the resonant frequency and cycle period gives a reliable condition to obtain plausible estimates of loses which is the key to determining scale factors from (6). Assessing the validity of the estimates is not simple as direct estimates of the subject vocal cords can not be obtained easily. Although the static mass could be estimated reasonably well using video images of the vocal folds of the subjects during phonation, it must be taken into account that static mass will differ from the dynamic equivalent masses of each vocal fold, as it can not be granted that the whole tissular structure of each vocal cord is equally involved in vibration (in other words, cords are not stand-alone free vibrating structures along their length, as they are connected to semi-rigid walls by the body conjunctive tissues, and thus the efficient contribution of the body to vibration is much less than expected). The estimations may be compared against values used by researchers working in vocal cord modeling. Story proposes masses of *0.058 g* (body), *0.082 g* (upper cover) and *0.094 g* (lower cover) and stiffness parameters of *112,250 dyn/cm*, *157,140 dyn/cm* and *203,660 dyn/cm* respectively for a *3-mass* model [9], each vocal cord assumed to be *1.5 cm* long and *0.3 cm* thick. Having into account that less than one third of the cord thickness is vibrating at the largest elongation, and this only along a part of the cord total length, the effective dynamic mass of such a cord would be about *0.0116 g* This evaluation of the dynamic mass is in the order of the estimations given in Table 1, much below the total static mass used in the model, estimated in *0.234 g*, (adding up the body and cover masses) which is clearly oversized attending to the cord dimensions proposed. The remaining cord structure (composed mainly by the bulk body tissue) less involved in strong vibration would contribute with more with elastic behavior rather than with dynamic mass. As such, part of the body structures are behaving more as elastic springs than as moving masses, and the estimations of the masses have to be substantially reduced. Estimates of cord stiffness should be proportionally reduced as well to keep resonant frequencies unaltered.

5. CONCLUSIONS

The research presented describes a possible approach to an interesting inverse problem: that of determining biometrical parameters from the glottal signals obtained from voice. The results show an order of magnitude disagreement with estimates classically used in vocal cord modeling, although a careful review of the concepts of static and dynamic mass explain the disagreement by the fact that only a small fraction of the total cord mass is involved in vibration. Having this into account the analysis of estimations for male and female adults and young reveal that the differences in dynamic masses involved in vibration are less relevant than cord tensions, and that estimated losses are in good agreement with classical model parameters. These results may open the path to use biomechanical unbalance and deviation as correlates of voice pathology.

6. ACKNOWLEDGMENTS

This research has been carried out entirely under the support of grants Nos. TIC2002-2273 and TIC2003-08756 from Programa de las Tecnologías de la Información y las Comunicaciones, Ministry of Education and Science, Spain.

7. REFERENCES

[1] Akande, O. O. and Murphy, P. J., "Estimation of the vocal tract transfer function with application to glottal wave analysis", *Speech Communication*, Vol. 46, No. 1, May 2005, pp. 1-13.

[2] Godino, J. I., Gomez, P., Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors. *IEEE Trans Biomed Eng.* Vol. 51, 2004, pp. 380-384.

[3] Gómez, P., Godino, J. I., Álvarez, A., Martínez, R., Nieto, V., Rodellar, V., "Evidence of Glottal Source Spectral Features found in Vocal Fold Dynamics", *Proc. of the ICASSP'05*, 2005, pp. 441-444.

[4] Gómez, P., Godino, J. I., Díaz, F., Álvarez, A., Martínez, R., Rodellar, V., "Biomechanical Parameter Fingerprint in the Mucosal Wave Power Spectral Density", *Proc. of the ICSLP'04*, 2004, pp. 842-845.

[5] Gómez, P., Lázaro, C., Fernández, R., Nieto, A., Godino, J. I., Martínez, R., Díaz, F., Álvarez, A., Murphy, K., Nieto, V., Rodellar, V., Fernández, F. J., "Using biomechanical parameter estimates in voice pathology detection", *Proc. of the 4th Int. Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications MABEVA'05*, Florence, Italy, October 29-31, 2005 (to appear).

[6] Gómez, P., Martínez, R., Díaz, F., Lázaro, C., Álvarez, A., Rodellar, V., Nieto, V., "Estimation of vocal cord biomechanical parameters by non-linear inverse filtering of voice", *Proc. of the 3rd Int. Conf. on Non-Linear Speech Processing NOLISP'05*, Barcelona, Spain, April 19-22 2005, pp. 174-183.

[7] McLaughlin, S., "Nonlinear Speech Synthesis", *Proc. of the EUSIPCO'02*, Vol III, 2002, pp. 211-218.

[8] Hadjitodorov S, Mitev P. A computer system for acoustic analysis of pathological voices and laryngeal diseases screening. *Med Eng Phys.* 2002;24: 419-429.

[9] Story, B. H., "An overview of the physiology, physics and modeling of the sound source for vowels", *Acoust. Sci. & Tech.*, Vol. 23, No. 4, 2002, pp. 195-206.

[10] Story, B. H., and Titze, I. R., "Voice simulation with a bodycover model of the vocal folds", *J. Acoust. Soc. Am.*, Vol. 97 (1995), pp. 1249–1260.